

## Article

# Driver Stress Detection from Physiological Signals by Virtual Reality Simulator

Nuria Mateos-García, Ana-Belén Gil-González <sup>\*,†</sup> , Ana Luis-Reboredo <sup>†</sup>  and Belén Pérez-Lancho <sup>†</sup> 

BISITE Research Group, University of Salamanca, 37007 Salamanca, Spain

\* Correspondence: abg@usal.es

† These authors contributed equally to this work.

**Abstract:** One of the many areas in which artificial intelligence (AI) techniques are used is the development of systems for the recognition of vital emotions to control human health and safety. This study used biometric sensors in a multimodal approach to capture signals in the recognition of stressful situations. The great advances in technology have allowed the development of portable devices capable of monitoring different physiological measures in an inexpensive, non-invasive, and efficient manner. Virtual reality (VR) has evolved to achieve a realistic immersive experience in different contexts. The combination of AI, signal acquisition devices, and VR makes it possible to generate useful knowledge even in challenging situations in daily life, such as when driving. The main goal of this work is to combine the use of sensors and the possibilities offered by VR for the creation of a system for recognizing stress during different driving situations in a vehicle. We investigated the feasibility of detecting stress in individuals using physiological signals collected using a photoplethysmography (PPG) sensor incorporated into a commonly used wristwatch. We developed an immersive environment based on VR to simulate experimental situations and collect information on the user's reactions through the detection of physiological signals. Data collected through sensors in the VR simulations are taken as input to several models previously trained by machine learning (ML) algorithms to obtain a system that performs driver stress detection and high-precision classification in real time.



**Citation:** Mateos-García, N.; Gil-González, A.-B.; Luis-Reboredo, A.; Pérez-Lancho, B. Driver Stress Detection from Physiological Signals by Virtual Reality Simulator. *Electronics* **2023**, *12*, 2179. <https://doi.org/10.3390/electronics12102179>

Academic Editor: Alberto Fernandez Hilario

Received: 3 March 2023

Revised: 28 April 2023

Accepted: 5 May 2023

Published: 10 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** photoplethysmography; stress recognition; virtual reality; intelligent mobile devices; machine learning

## 1. Introduction

Affective computing is an interdisciplinary field that focuses on the development of systems and devices that can recognize, interpret, process, and simulate human affect. An increasing number of researchers have conducted studies on affective computing in general, and emotion recognition in particular, making it an emerging and promising area of research [1]. Systems have been developed that can adapt to the emotional state of users to enhance learning and experience [2–5].

Recent statistical analyses have revealed that human behavior is one of the main causes of road accidents [6]. Research has shown that drivers' states of mind and different situations occurring during a road journey, such as fatigue, traffic jams, junctions, and traffic lights, affect the health and stress of drivers and passengers [7–10]. Currently, it is known that stress is one of the most common mental health problems suffered by the population and that it has very detrimental effects on health [11]; therefore, its detection can lead to stress mitigation. For this reason, monitoring some physiological variables of the user in real time can be key to detecting their behaviors [12–14].

Technological evolution has led to major developments in wearable sensors capable of capturing various human physiological measurements in a cost-effective, non-invasive,

and efficient manner, with similar reliability to other traditional methods, such as electroencephalogram (EEG) [15], where monitoring with these devices is activity-restricting, costly, and cumbersome. Therefore, the feasibility of detecting an individual's stress state using a PPG sensor embedded in a wristwatch, which is widely used today, was investigated. Different devices have been used to perform multimodal approaches.

The main goal of this work is to combine the use of these sensors with the possibilities offered by VR to create a system for recognizing the stress of an individual during different driving situations in a vehicle. The combination of visual and audio stimuli and the possibility of interaction and total immersion of the user in a three-dimensional environment enhance the realism and impact of emotional stimuli [16–18]. Thus, an innovative method of learning and stimulus induction is proposed, which safely circumvents space and time constraints through VR head-mounted display (HMD), that is a 3D visualization device. This technology allows users to immerse themselves in purpose-built virtual spaces, recreate scenarios that are very difficult or even impossible in real life, such as, in this case, an accident, and users can even interact and manipulate objects freely and safely, using controllers or other technologies such as eye-tracking or motion sensors [17]. The use of VR technology to elicit emotions has been shown to increase and stimulate the physiological responses of subjects as the results obtained consisted of an increase in the mean electrodermal activity (EDA) and heart rate (HR) compared to a standard or non-immersive (2D) simulation during a driving simulation [19].

A commercial smartwatch was used to acquire PPG and respiration data, which has been found to be effective for the recognition of basic emotional states [9,20]. The smartwatch collects information from HR, oxygen saturation (OS), and heart rate variability (HRV), from which time and frequency features are extracted to select the most suitable ones for the emotional classification algorithm. The Leap Motion hand-tracking device was used to observe the participants' movements during the simulation.

Data collected from sensors are graphed, integrated into a broader context mobile system, and stored on the smartphone in a format accessible from any computer, in this case, comma separated value(s) (CSV), to perform real-time monitoring and visualization of selected features, as in [21], while processing and performing classification tasks. The first consists of processing the signals and extracting relevant features for classifiers, tasks, and useful knowledge that can be obtained through big-data-mining techniques. The second phase involves the use of ML technologies, which are widely and successfully used to solve real-world questions. An increasing number of researchers are tackling decision-making and classification problems with the help of human-brain-inspired, bio-inspired algorithms, namely neural networks [2,3,5,22,23], because they are very flexible algorithms that are applicable to diverse sources of different types.

The remainder of this paper is organized as follows. Section 2 compiles related work, issues, and trends in which physiological signals and immersive methods enable automatic recognition of emotions. Section 2.2 focuses on stress detection, specifying the scope of this work. Section 3 describes the fundamentals that allow the detection of emotional states in humans through the automatic procedures of artificial intelligence (AI) with the support of sensors. Section 4 presents the methodology used for the analysis of the recognition of stress in drivers using immersive methods. Section 5 presents a case study detailing the immersive process and operativity of the models with the collected data. Section 6 shows the results obtained. The paper concludes in Section 7, which includes a discussion of the results obtained, conclusions, and possible future lines of research.

## 2. State of the Art

The use of physiological signals for affective computing is very interesting since they have numerous advantages. One of the keys is objectivity, since these types of signals are regulated from the central nervous system, directly reflecting mental activities, without being able to consciously control, hide, or alter them, which makes them very effective signals for recognition of emotions in multiple situations.

On the other hand, the growing popularity and evolution of wearable devices, which are becoming more powerful and more widely used, make it possible to easily acquire physiological signals from users at low cost and with a long lifetime, capturing their behaviour independently of the posture adopted by the participant. The user's context also affects the individual, whether it is location, time, culture or individual characteristics, so that emotional states vary from person to person [24].

Facial recognition (FR), electroencephalogram (EEG), and speech recognition (SR), along with heart-related methods such as blood volume pulse (BVP) or electrocardiogram (ECG) allow for independent measurements as they can recognize emotions in the valence and arousal dimensions [25–27].

The galvanic skin response (GSR) or electrodermal activity (EDA) limits the recognition of emotions to certain states such as fear [18], concentration, depression [5] or stress [28], but makes it difficult to detect broader spectrums. Additionally, together with the skin conductivity (SC), they only detect the arousal level. The opposite case occurs in the electromyography (EMG), which only classifies the valence level [17,25].

The breakdown of data in state-of-the-art studies is not entirely transparent, making it difficult to compare measurement methods. Table 1 offers an overview of the modality, the advantages, the disadvantages, and the areas of the methods that have been used.

**Table 1.** Comparison and study of the measurement methods regarding the benefits, limitations, and areas of application.

|     | Benefits                                                                         | Limitations                                                                                                 | Application Area                                                       | Ref.                 |
|-----|----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|----------------------|
| PPG | Continuous monitoring, freedom of movement, low cost.                            | Sweat affects measurements.                                                                                 | Anywhere.                                                              | [29]                 |
| EEG | Allows measurements on disabled people.                                          | Complex installation, high maintenance, limited movement.                                                   | Laboratory conditions.                                                 | [4]<br>[23]          |
| ECG | Mobile measurements (smart devices) and relevant data acquisition.               | Increased accuracy in stationary measurement and motion artefacts in mobile systems.                        | Laboratory conditions, daily use, sports activities.                   | [30]<br>[31]         |
| EMG | Allows measurements in people with psychological disorders.                      | Single measurement of valence, difficult installation, amplitudes vary depending on the measurement chosen. | Laboratory conditions.                                                 | [25]<br>[32]<br>[17] |
| FR  | Multiple person tracking, contactless.                                           | Requires a front camera, fakeable.                                                                          | Laboratory conditions, public spaces, workplace, home automation.      | [33]<br>[34]         |
| SR  | Non-contact, casual measurement.                                                 | Requires a microphone, communication necessary, prone to ambient noise, counterfeitable.                    | Wide field of application, assistants or smart calls.                  | [12]<br>[35]         |
| BVP | Versatile due to its small sensors, evaluation of health parameters.             | Prone to artifacts in some areas of application (sports movements).                                         | Laboratory conditions, daily use, sporting activities (smart devices). | [32]<br>[13]         |
| EDA | Successful stress indicator.                                                     | Unique excitation measurement, influence by temperature, need for referencing and calibration.              | Laboratory conditions, daily activities (Empatica E4).                 | [28]<br>[5]          |
| RP  | Easy installation, indicates emotional states (fear, depression, concentration). | Difficult to distinguish from a broad spectrum of emotions.                                                 | Wide field of application.                                             | [31]<br>[36]         |
| SC  | Multiple ways of data acquisition (video, sensors, infrared).                    | Unique arousal measurement, dependent on external temperature, slow indicator for emotional states.         | Laboratory conditions, public spaces, workplace, home domotics.        | [22]<br>[24]         |

For this reason, many studies recommend combining data from methods based on physiological signals to recognize the wide range of emotions. This multimodal approach causes more complex data processing, but increases the accuracy of the classification.

Specifically, after reviewing up to 47 studies that compile ECG, EDA and accelerometer (ACC) signals from 2015 to the present, they have proven to be valid for recognizing the 8 basic emotions published by Ekman in recognizing stress, levels of stress, relaxed state, or neutral state for the detection of positive and negative emotions and different quadrants in the valence and arousal space. However, for stress detection, many studies have shown that the PPG sensor, capable of providing HRV information, achieves good classification accuracy.

Emotional elicitation based on visual stimuli is widely used because it increases emotional responses [23]. Images, sounds, music, or videos are one of the main channels of information in interpersonal communication [37]. Writing, tactile detection, or games have also been used to induce emotional states. On the other hand, video games are interactive software capable of arousing different emotions in players [38]. It is crucial to accurately perform each stage of the game to induce different types of emotions in users [26]. The main theories applied during the game design process are based on the observation of the behavior of the players and/or the experience of the authors [26].

Virtual reality (VR) can be understood as an extension of the previously mentioned stimuli, sounds, images, and games, with the added value of immersion and greater reliability in the obtaining process [16]. This affects the behavior of the users, eliminating the passive role they take in the face of limited interaction stimuli, which results in a greater emotional or bodily manifestation [38].

Combining the use of VR glasses adds an additional difficulty since the helmet or HMD hides a part of the face, preventing the detection of facial features, which have proven to be effective solutions in the classification of emotions [27,33], although they can be falsified [39], but in other work this may be beneficial as the HMD can be used to implant eye sensing systems [17] that allow interpretation of eye movements, such as the detection of saccadic movements and pupil dilation.

In [18], a virtual environment is proposed to investigate the mechanisms of emotional learning to improve treatments that address the severe effects of anxiety and fear disorders among humans. Verbal and physiological cues and behavioral data are collected from various female and male participants with low and high social anxiety who performed the virtual simulation consisting of approaching virtual agents of both genders using a *joystick*.

Regarding the algorithms, the most used by the authors are artificial neural networks (ANN), convolutional neural networks (CNN), or dynamic neural networks (DNN), especially when the emotion activation stimulus is images and the EEG response is evaluated. For multisensor devices with ECG, GSR, blood pressure variability (BPV), EDA signals, the most used have been decision tree (DT), K-nearest neighbor (KNN), random forest (RF), and support vector machines (SVM), and with these last two, the highest percentages in classification have been achieved. Multilayer perceptron (MP), Hoeffding tree (HT), and naïve Bayes (NB) are also used in a minority of cases.

In the following sections, the works related (1) to the stimuli used to induce the emotional state and (2) the ML models used for the classification of stress or other emotional states by means of portable devices and signals that can be captured by PPG are compiled.

### 2.1. Automatic Emotion Recognition in Immersive Media

One of the works related to the classification of emotions in immersive media [13] focuses on inducing and detecting a person's anger while driving. Three road simulation scenarios are developed to cause driver anger: waiting for the user when the traffic light is red, a traffic jam, and crossing a vehicle. The driving simulator consists of a vehicle, a 180° screen, and five networked computers, with the addition of a steering wheel, a gearshift lever and accelerator, and clutch and brake pedals. The virtual scenario realized was an effective method of provoking anger. During the simulation, biological and brain signal data were collected from 15 licensed participants with previous driving experience. The recordings were made with an EEG computer and the Biograph Infiniti system, a software that captures, analyzes and graphs psychophysiology and biofeedback data. For

classification, the collected data were randomly divided into two sets, with 80% being the training group and the remaining 20% being the test aggregate. A hidden naïve Bayes (HNB) classifier was used, obtaining an accuracy of 87% in the scenario of another vehicle crossing and an evaluation metric of area under the receiver operating characteristic (ROC) curve (AUC) of 0.98, so the performance of the classifier is good. In this work they highlight the importance of simulation since it allows us to perform this experiment avoiding possible accidents that could occur in real situations.

The paper [16] proposes a user-independent emotion recognition system that collects multimodal data from ECG, EDA, BVP, respiration, and HR sensors during an elicitation protocol based on VR and records them in a computer application. A device placed on the arm and another on the chest of the participants allows for the collection of the signals. The virtual scenario consists of a series of three-dimensional videos that are played through HP's mixed reality goggles and headset to a total of 23 participants with no history of psychological or neurological conditions and taking no medication in the days prior to the experiment. The extracted physiological features are used as input to the SVM-based emotion recognizer, which uses a public database of immersive VR videos with ratings in arousal and valence [40]. An accuracy of 69.13% is obtained for arousal and 67.75% for valence, and an accuracy of 85.3% is obtained for the recognition of three emotions, amusement, interest and relaxation, by using ECG, EDA, temperature and RP.

One of the main advantages of VR is that the damage that can be caused in real situations is eliminated in addition to its applicability in any field. The literature review shows that, despite its great potential for immersion and stimulation of sensations, VR remains unexplored within the state of the art. The breakthrough in technology overcomes problems previously encountered, such as the need to connect the HMD to a computer, which restricts the play area and requires the experiment to be conducted in a laboratory context. These advances, together with modern portable data acquisition devices that are small in size, non-intrusive, and provide freedom of movement, and the great current impact of the IA and ML algorithms allow us to unite both technologies in trend for the development of the software proposed in this work, focused on a very important area, as traffic accidents are one of the major causes of mortality in the world.

## 2.2. Stress Detection: Related Work

Negative emotions can induce weaknesses and cause risks to the immune system [41]. Emotional sensing systems offer the potential to identify them and be able to address the causal factors and are useful in applications in multiple domains, especially those focused on health, such as stress [16,38,42] or other types of emotions [3,5,31,43,44], where physical and mental states can be monitored in real time [2,24,36] and can act accordingly [28] as intelligent assistants [21,45]; in environmental assisted living [27,33]; in the industries of games [26], robots [46,47], domotics [48], marketing, or recommendations [4,34,37,49]; and in the study of social behavior [23,30,34], authentication and security [18,39,48], or education [49], among others.

To detect emotions, systems have been developed that use different channels of human emotional expression, such as tone of voice, facial expression, body posture, behavior, or handwriting. Although most emotion recognition studies fuse data collected from multiple sensors (audio, video, computer records, facial expression, posture, and different types of physiological characteristics), many of the proposed methods violate privacy and security, such as video or speech recordings or keystroke logging, and others limit movement and may even generate stress as they are unusual methods.

The trend of globalizing many smart sensors in the vision of the Internet of Things or IoT, accompanied by the growing evolution and use of electronic devices, cell phones, tablets, wearables such as watches (smartwatch) and smart bracelets, clothes or sneakers that integrate a global positioning system (GPS), headbands, and other types of wearable devices that are incorporated into some part of the human body, have enabled efficient meth-

ods of modeling and processing contextual information, interacting continuously with the subject and in conjunction with other devices to perform some specific function [2,41,50].

In addition, smart wearable sensors enable the implementation of emotion detection methods by using integrated ML algorithms [51,52], which are able to combine multiple biomarkers of an emotion type extracted from different physiological signals and are thus currently being used for emotional state recognition.

Devices such as watches are common in the population. Some studies have demonstrated the feasibility of detecting emotional states, especially stress, through such devices. In [53] they achieved an accuracy of 81.2% in detecting three emotional states: neutral, anger, and happiness using SVM and using DT and RF they achieved an accuracy of 83.3% and 73.8% while watching video clips developed to elicit happiness and anger. In [54], with HRV signals, it achieves an accuracy of 84%, a value of 78% in the AUC metric and 0.56% in F1 to detect whether a subject is stressed or not stressed, and in [20], the authors classify various emotions with an algorithm based on SVM with an accuracy of 76%.

One of the main challenges of emotion recognition from physiological signals is the acquisition of a representative sample of data. There are several public physiological signal databases built for the purpose of classifying previously labeled emotions.

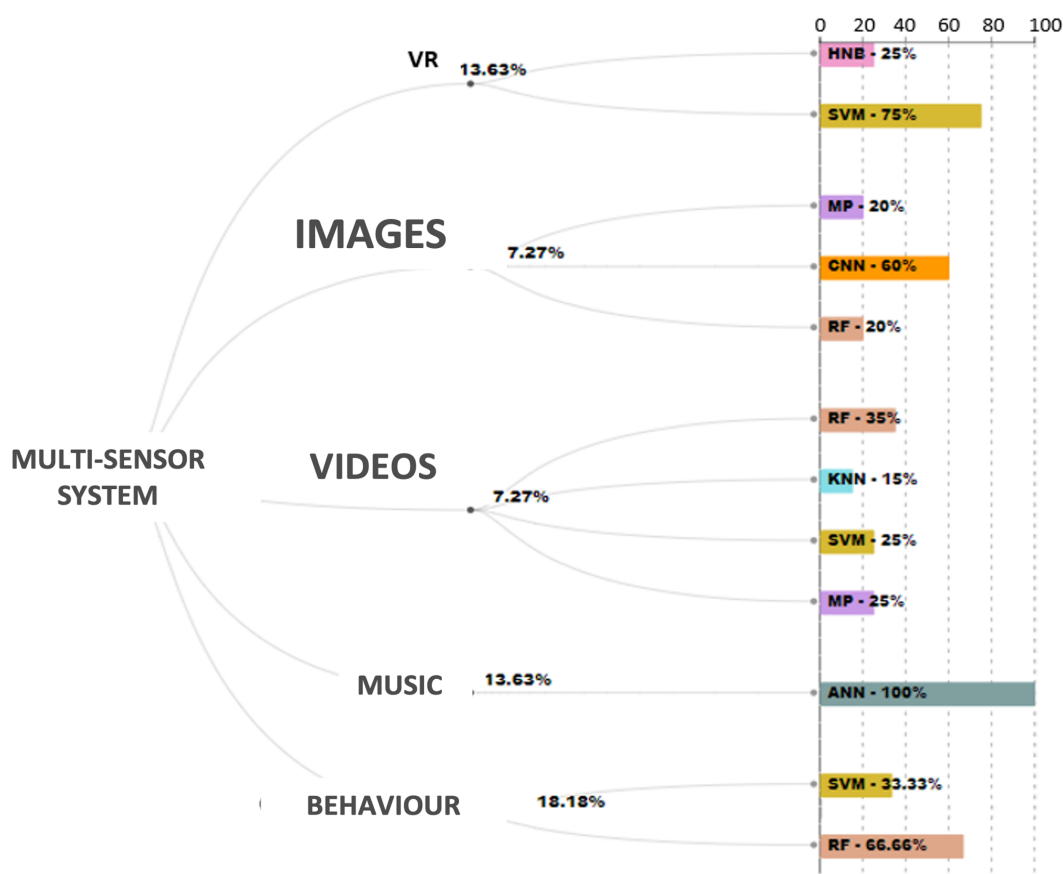
Table 2 collects the datasets most related to the case study. They collected signals while inducing different stimuli to elicit different solutions in a number (n) of participants. Many widely used databases for emotion recognition had to be discarded as they required information from EEG signals, such as MAHNOB-HCI, ASCERTAIN, MPED, AMIGOS, DEAP, SLADE, DECAF, DREAMER, or CLAS.

**Table 2.** Review of available related databases.

| Name                                                  | Year | Signal                               | Stimulus                                                         | N  | Output                                                                 | Ref. |
|-------------------------------------------------------|------|--------------------------------------|------------------------------------------------------------------|----|------------------------------------------------------------------------|------|
| Stress Recognition in Automobile Drivers              | 2005 | ECG, GSR, HR, and breathing.         | Driving.                                                         | 16 | High, medium, and low stress.                                          | [55] |
| Eight Emotion Sentics Data                            | 2000 | ECG, EMG, EDA, HR, BVP.              | Driving.                                                         | 25 | Neutral, anger, hate, sorrow, love, romantic love, joy, and reverence. | [56] |
| Affective ROAD                                        | 2008 | EDA, HR, GSR, HR, and hand movement. | Driving.                                                         | 12 | Degree of stress.                                                      | [9]  |
| WESAD                                                 | 2018 | ECG, ACC, EDA, HR, EMG.              | Reading, videos, and mental arithmetic.                          | 15 | Neutral, fun, and stress.                                              | [42] |
| MMOD COG                                              | 2019 | ECG, EDA, HRV, and voice.            | Reading, testing, and resting.                                   | 40 | Level of cognitive load.                                               | [57] |
| SWELL KV                                              | 2018 | EDA, HRV.                            | Neutral, interruptions, pressure, relaxation.                    | 25 | Emotion, stress, and relaxation.                                       | [58] |
| Biometrics for stress monitoring                      | 2019 | EDA and HRV of SWELL and WESAD       | Neutral, interruptions, pressure, relaxation.                    | 25 | Stress, fun and neutral types.                                         | [59] |
| Non-EEG dataset for Assessment of Neurological Status | 2017 | EDA, SPO2, HR.                       | Relaxation, physical stress, cognitive stress, emotional stress. | 20 | Neurological status and stress.                                        | [15] |

Therefore, most authors use HR data to recognize some kind of emotion. In recent years, systems capable of recognizing emotional states solely with the information provided by the PPG signals have been developed, as in [53], where the state of happiness is detected with an accuracy of 80.38% or in [60] with an accuracy higher than 87% in the detection of happiness, sadness, or neutrality.

Emotional responses must be elicited using stimuli in order to measure them. The stimulus must elicit a specific emotion in the subject. Different scenarios have been used to elicit emotions, and the most commonly used are shown in Figure 1, together with the algorithms used in the classification.



**Figure 1.** Dendrogram with the percentage of stimulus induction according to the most used stimuli and the most used algorithms in each case.

More than half of the works found in the literature use film clips or images as they are a rich source of discrete emotions (joy, love, fear, anger, etc.) with representations of events and scenes from everyday life and images the same as videos, but with a shorter stimulation time. Music has also been used to develop emotions over time in a simple way, although it is dependent on the participant's musical taste, which may influence emotions. VR, which combines several stimuli such as audio, visual, and events or interaction results, is a booming technology as a scenario for emotion induction, especially that focused on evaluation and intervention processes in people with some type of disability, since other authors have shown that greater benefits can be obtained in treatments based on this technology [61,62], where SVM is the most used for classification. Although only 13.63% of the scientific literature has used VR for the elicitation of emotional states (Figure 1), this technology has been used in recent decades in several papers as a tool to assess people's driving behavior and its correlation with physical responses [19,29]. Several studies have evaluated the physiological response to driving using a virtual driving simulator, collecting the signals from HR and HRV [29], EDA [10], EMG [8], and EEG [63]. In all of them, the VR potentiates the physiological response of the subjects during the simulation. Emotion eliciting conditions include physical fatigue [64,65], inappropriate behavior of other drivers [13], mental workload [7], differences in braking signals [63], a crash [8,29], or during certain emergency maneuvers [19].

In addition, nowadays, modern HMD have greatly improved technology and are capable of eliciting a greater subjective sense of presence [19], which should be reflected physiologically while driving in the proposed virtual simulator.

Table 3 gathers the works with the best results in stress classification, only with input variables related to the signal PPG or ECG captured through portable devices. Models

are built that obtain accuracies better than 71% in stress classification. The best results are obtained with SVM. Mostly, *wearables* are used for signal acquisition.

**Table 3.** Related studies on the classification of emotions through portable devices.

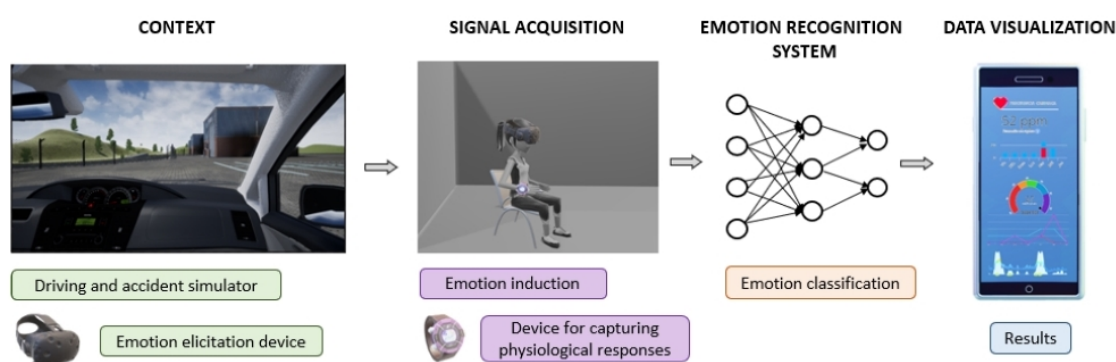
|      | Stress    |           |        | Stimuli | Device    | Sensor | Accuracy       |               |               |               |
|------|-----------|-----------|--------|---------|-----------|--------|----------------|---------------|---------------|---------------|
| [53] | Happiness | Neutral   | Stress | Video   | Watch     | PPG    | SVM:<br>81.2%  | DT:<br>83.3%  | RF:<br>73.8%  |               |
| [54] | Stress    | No stress |        | Daily   | Mobile    | ECG    |                | SVM:<br>84%   |               |               |
| [20] | Stress    | No stress |        | Video   | Wristband | PPG    | SVM:<br>75.56% | NN:<br>73.92% | RF:<br>73.61% | NB:<br>71.38% |

### 2.3. Conclusions and Proposal

Although there are studies that analyze physiological responses during vehicle driving, the state of the art, collected in Section 2, has the following limitations: (1) few studies analyze physiological responses in driving using an affective approach; (2) a minimal number of papers use only the PPG sensor for signal capture in emotion recognition; (3) there are no validated driving simulator datasets that include stimuli with different arousal and valence levels; and (4) there are no papers on automatic emotion recognition during driving simulation by collecting the physiological signals induced by VR and using machine learning algorithms.

In this work, as shown in Figure 2, we propose a system that attempts to take advantage of the technological evolution in the PPG sensors of common and low-cost portable devices to acquire information about the HRV and feed the ML model in such a way that it classifies the stress state of the user while performing a driving simulation in VR, allowing scenarios impossible or dangerous to be recreated in real life and is able to obtain very valuable information about the human behavior and emotional responses of individuals during driving and their reactions to accidents.

The stimuli were designed to induce different emotional states during possible hazards that may occur while driving a vehicle, such as carelessness or brake failure, using an HMD device. The physiological signal acquisition device used was a Garmin Forerunner 235. The results were displayed on a mobile device.



**Figure 2.** Structure of the proposed system.

### 3. Conceptual Foundations of Emotion Extraction

Emotion is a type of conscious or unconscious feeling expressed through various biological and physical reactions that occur in response to certain external or internal stimuli [3].

The study of emotions and their influence on physiological behavior is an important topic. The mental and physical health of people is influenced by emotional changes as these changes can affect the presence of variations in respiratory activity or cardiovascular

behavior [31]. In many cases, people are not aware of their own condition, and the possible influence this can have on the proper performance of different systems, such as the cardiac system [31,41]. Studies have shown that there is a relationship between physiological signals and the arousal and valence dimensions of a felt emotion, terms detailed in Section 3.1, allowing emotions to be classified as positive or negative according to the circumplex model of emotions, as described in Section 3.1, and providing pleasant, dangerous, or unpleasant situations. Multiple studies have examined the use of multimodal systems to capture physiological signals and map them to emotional states [30].

The goal of affective computing is to enable recognition of human emotional states using automatic procedures [50].

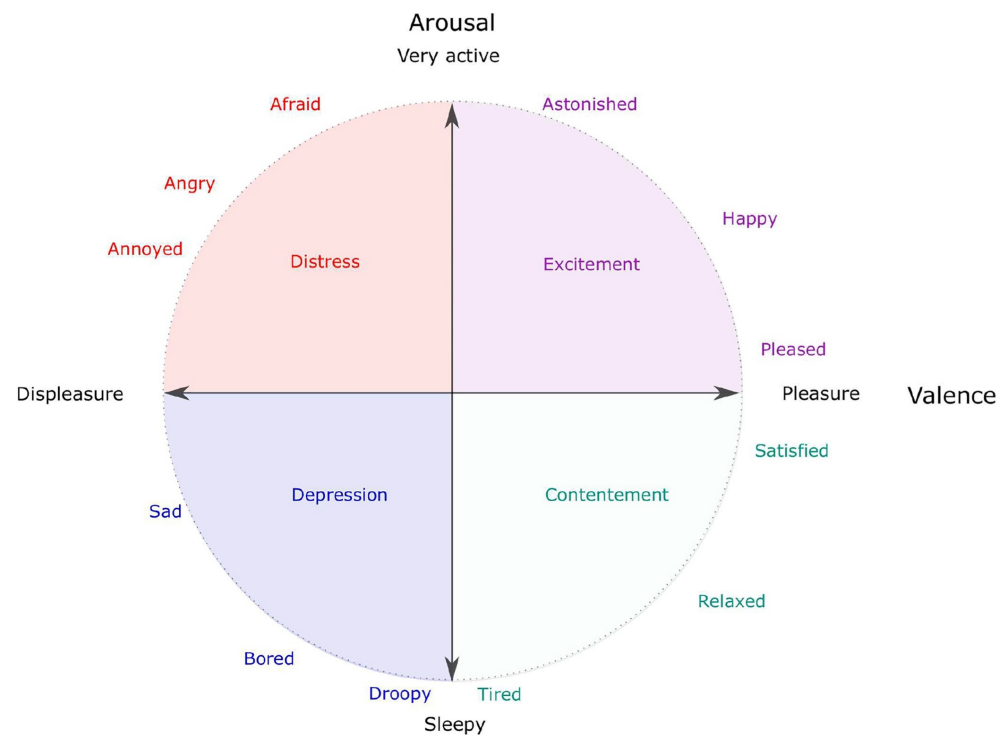
1. Participants: the method for building the emotion classification can be divided into subject-dependent and subject-independent models. The former means that the model is built for each new user; therefore, the accuracy is higher than in the second case, where the model is formed for the entire database.
2. Emotion model: emotional states to be perceived.
3. Stimuli: way to induce emotions. The most commonly used are images, videos, sounds, and even natural stimuli [66] such as noise, temperature, or interaction with other people or animals [30].
4. Data acquisition: systems used to capture human responses to stimuli. In the Section 3.2, the most commonly used existing methods are described in Section 3.2.
5. Data processing: This stage consisted of two phases.
  - Signal processing, filtering the data for meaningful biomarkers [39].
  - The extraction of potential signal features, where the previous phase data is segmented into time windows (which may overlap) and from which different relevant features, such as time and frequency domain, are extracted [24].
6. Classifier or learning model: In recent research, various ML algorithms have been trained at this stage, and the models have been evaluated. First, the most important features were selected by cross-validation on the training set and then to ensure model generalization on an independent test set. Section 3.3 presents the ML algorithms used for emotion-state recognition.

### 3.1. Emotion Models

The number of emotion categories has been a question mark since the inception of the study of emotions by psychologists, although there are two key methods: basic emotion theory, which labels emotions into discrete categories, and multidimensional theory, which classifies emotions into multiple [4] dimensions.

Discrete emotion theory describes basic emotions as discrete as it considers that they can be distinguished by the biological processes or facial expressions of individuals. One of the most important studies was conducted by Ekman in 1972, who considered six basic emotions: joy, anger, fear, sadness, surprise, and disgust [12], each with its own characteristics that allow it to be differentiated from others. Later, in 1990, he expanded the list and included a wider range of secondary positive and negative emotions, many of which were combinations of two or more basic emotions [3].

Russell proposed the circumplex model of emotion, in which emotional states are represented in a space of two basic dimensions: the horizontal dimension or valence, which corresponds to positive or negative emotions, and the vertical dimension, which corresponds to the degree of arousal and relaxation [32] such that the center of the circle represents a neutral valence and a medium level of arousal. According to this concept, emotions can be defined in the regions within the emotional plane as a combination of valence and arousal [67], as shown in Figure 3. This model is one of the most commonly used for testing stimuli and emotional states.



**Figure 3.** Russell's circumplex affect model.

Generally, arousal denotes emotional intensity and valence, the type of emotion (multiple levels between sad and happy), and dimension values are discrete, for example, low or high binary states (forming four quadrants in two-dimensional space) or three values: low, medium, or neutral and high [3].

### 3.2. Data Acquisition

At present, there are simple, low-cost and portable data acquisition and processing systems. Among the wide variety of existing methods used to characterize changes in physiological activity due to emotional influences, the most commonly used are as follows:

- Electroencephalogram (EEG) is the most direct method because it measures the electrical activity of the brain. This allows the discrimination of different levels of excitation or stimuli with positive or negative emotional valence [3].
- Electromyography (EMG) detects the electrical activity of facial muscles involved in the expression of emotional states, and the data obtained are valid for the identification of negative or positive emotions [32].
- Blood pressure variability (BVP) is highly associated with cardiovascular disease [13] and measures the flow of blood circulating through the veins, allowing calculation of the HR. High levels of BVP can occur in users when they are in a state of stress or anger, whereas low BVP values can indicate states of sadness or relaxation [13].
- Galvanic skin response (GSR) or electrodermal activity (EDA) indicates variations in the electrical conductance of the skin caused by the action of sweat glands in response to emotional changes. It is used as an indicator of arousal or stress [36,42].
- Temperature is regulated by the sympathetic nervous system and depends on the underlying blood flow, which is reduced when muscle fibers are activated. Temperature changes during the expression of certain emotions such as fear or anger [18]. This signal varies and can be influenced by sweating, changes in the body, or environmental conditions such as humidity or temperature. It is sometimes used to differentiate between conflict and non-conflict situations and is therefore related to the arousal dimension [18].

- The Heart rate (HR) is the number of heartbeats elapsed in one minute. HR values vary with emotional states. Heart activity can be detected using several methods.
  - The electrocardiogram (ECG) measures and detects electrical signals in the arms or chest. A low HRV value may indicate a relaxed state, whereas an increase in HRV may indicate a state of distress or frustration [32]. With the ECG signal, other physiological signals in addition to HR can be calculated, such as the inter-beat interval (IBI), HRV and respiratory arrhythmia [12]. An increase in IBI may indicate stress or frustration [32].
  - Photoplethysmography (PPG) uses light to measure changes in light absorption in the skin. It is a very often used method to measure HR in a non-invasive, portable, and low-cost manner, but with good results as it can even replace ECG recordings for the extraction of HRV signals, especially in healthy individuals. The PPG uses an infrared light to measure the volumetric variations of blood circulation—very important information in the subject of health. The operating process is based on the projection of light onto the skin through the device's green LEDs, which reach the blood circulating through the individual, bounce back, and a photodiode placed under the device captures that light. The greater the amount of blood circulating through the blood vessels, the greater is the reflected light, and vice versa, thus knowing the HR.
  - Phonocardiogram (PCG) uses a microphone to acquire the signal or heart sounds.
- The respiratory pattern (RP) measures the speed, frequency, and depth of breathing, which can reflect fitness, health, and emotions. Deep rapid breathing may indicate a state of amusement or anger, while shallow rapid breathing may indicate a state of tension caused by fear, panic, or concentration. On the other hand, deep slow breathing may mean that the user is in a relaxed state, while shallow slow breathing may indicate a depressed state [12].

However, a context-aware system (CAS) is any information that can be used to characterize the situation of an entity [68]. This broad definition allows a variety of factors to be considered when determining the context.

- Physical, collected using device sensors, such as user location, activity log, calorie consumption, sleep status, or luminance.
  - Environmental, obtained through software services, such as weather or traffic at the user's location.
  - Organizational, stored on an electronic device, such as messages or calendar events.
- Based on this, a high-level context can be generated, interpreted, and recognized.

### 3.3. Classifiers and Machine Learning Algorithms

After the extraction of the relevant features for the possible differentiation of emotional states, the next step is to use them to train a model that allows the automatic classification of emotions. The aim of this study was to automatically detect the stress of an individual while driving.

Because training data are labeled, ML-supervised algorithms have been used to try several models such as multilayer perceptron (MP), Gaussian naïve Bayes (NB), K-nearest neighbors (KNN), bagging, random forest (RF), and support vector machines (SVN), as detailed in Section 4.3.

### Model Evaluation Metrics

Once algorithms for the construction of different classifier models have been applied, it is necessary to evaluate their classification performance. For this purpose, the most common metrics used are: precision, accuracy, F1 Score, recall, ROC, and confusion matrix.

The true positive (TP) is the number of positives correctly classified by the model as positive. True negative (TN) is the number of negatives correctly classified as negative by the model. Conversely, false positive (FP) is the number of negatives that were incorrectly

classified by the model as positive, and false negative (FN) is the number of positives that were incorrectly classified by the model as negative.

Precision returns the proportion of correct class identifications, as shown in Equation (1).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

Recall or sensitivity is similar to precision but returns the proportion or amount of TP that the model correctly identifies, as seen in Equation (2).

$$\text{Recall/Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

The specificity returns the proportion of TN rate, as in Equation (3).

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

Accuracy measures the percentage of cases in which the model is correct. It tends to perform poorly when there is class imbalance. Equation (4) contains the following formula for this metric:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

The F1 value combines the recall and precision measurements into a single value, as shown in Equation (5).

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

The confusion matrix makes it possible to visualize the number of instances of each class classified as correct or incorrect in a simple manner (Table 4).

**Table 4.** Confusion matrix.

|                     |           | FORECAST (Predicted Class) |           |
|---------------------|-----------|----------------------------|-----------|
|                     |           | Positives                  | Negatives |
| FACT (Actual class) | Positives | TP                         | FN        |
|                     | Negatives | FP                         | TN        |

Once the machine-learning model is built, it is necessary to determine the effectiveness of our model based on metrics and datasets. The ROC (receiver operating characteristics) curve is generated by calculating and plotting the true positive rate (TPR) against the false positive rate (FPR) to show the performance of a classification model at different thresholds, as shown in Figure 4. AUC (area under the curve) represents the two-dimensional area below the curve and indicates the model's ability to distinguish between classes. The higher the AUC, the better the model distinguished between the two classes. The figure below shows what the ROC curves would look like for three different hypothetical classifiers. The perfectly fit model had an AUC of approximately 1.

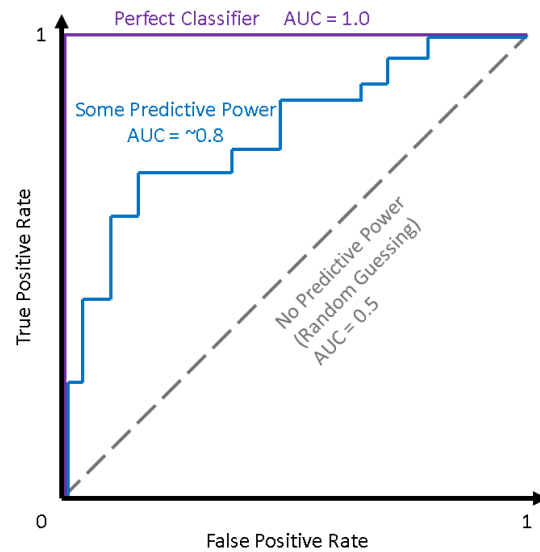


Figure 4. AUC–ROC Curve.

#### 4. Methodology

To achieve the objectives of this work and develop an automatic emotion recognition system, several procedures were carried out: (1) the search and selection of a relevant database, (2) the development of an application for the induction of emotional states to users, (3) the choice of devices capable of capturing the signals emitted in response to the stimuli, (4) the introduction of the information collected by the sensors as input to the neural network, (5) the application of different ML algorithms to compare the results and validate them according to the different existing metrics, and (6) the development of a mobile application that integrates the classifier system or neural network to determine the user’s emotional state and visualize relevant features for the user.

##### 4.1. Dataset

The dataset used was created in 2007 and is titled “*Stress Recognition in Automobile Drivers*”. It is available free of charge on the PysioNet platform (PhysioNet: research resource for complex physiological signals. Available at: <https://physionet.org/content/drivedb/1.0.0/>, accessed on 8 May 2023) and in the Google Cloud (<https://console.cloud.google.com/storage/browser/drivedb-1.0.0.physionet.org>, accessed on 8 May 2023).

For the selection of this set, an exhaustive search of all the databases related to automatic emotion recognition was carried out, as shown in Table 2, where many were discarded for having characteristics or data collected from other types of sensors or for merely inducing emotions without determining the elicited emotional states. In addition, a rubric was used to establish certain evaluation criteria and score the different sources. These criteria include that the database has a standardized identification, associates an accession number identifying the dataset, incorporates metadata, describes the experimental methods used to generate the data, is free of a university licence, and provides and describes how to cite the dataset.

PysioNet’s dataset scored the highest and was chosen because of its similarity of intent to the study’s objectives. The database collects information from sensors ECG, GSR, and respiration during the driving of a real vehicle to determine the driver’s stress level [55]. In addition, the signals were monitored in a relatively stationary position, as the subject was seated, so the signals could be clearer and more similar to those collected in the experiment proposed in this work. In the tests conducted in [55], drivers drive along a set route through open roads in the Boston metropolitan area; that is, they drive in normal or stressful environments with red lights, traffic jams, etc. Data were collected from 24 subjects

for at least 50 min, and 5-min data intervals during rest, highway driving, and city driving were performed to distinguish various stress levels with greater than 97% accuracy, as well as to compare continuous features calculated at 1-second intervals, where the results showed that for most drivers, HR correlates very closely with their stress level.

#### 4.2. Stress Classifier Model

From the initial dataset, several models are created and trained and optimized through the techniques of ML so that the system can learn from the data (the emotional truth labels are used to train the model) and apply it to a new dataset collected with the device. Information and data were collected during the experiment in real scenarios, that is, driving and accident simulators, to determine whether the user is stressed or not and to monitor the different characteristics [69].

##### 4.2.1. Preprocessing of Data

The first step was an initial exploration of the car driver stress recognition data to determine what characteristics or attributes were handled in the database, what type of data they were, and a statistical description. The dataset had 23 attributes and 4129 instances. All these correspond to the signals of ECG, EMG, and GSR. Table 5 lists the features that have not been used for the realization of the model, the reason behind the decision, and those that have been used for the construction of the classifier.

**Table 5.** Study of the characteristics of the dataset.

| Deleted Features                          |                                                            | Valid Features                                             |                                                                        |
|-------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------------------|
| <b>footGSR</b><br><b>handGSR</b>          | Sensor not available                                       | <b>HR</b>                                                  | Instantaneous heart rate                                               |
| <b>EMG</b><br><b>ECG</b>                  | Difficulty of acquisition                                  | <b>RESP</b>                                                | Number of breaths per minute                                           |
| <b>ULF</b><br><b>VLF</b>                  | Media = 0<br>Long recordings                               | <b>Seconds</b><br><b>newtime</b>                           | Time between wave intervals                                            |
| <b>LF_HF</b>                              | Missing data                                               | <b>LF</b><br><b>HF</b>                                     | Frequency domain                                                       |
| <b>interval in seconds</b><br><b>time</b> | Correlation 1 with AVNN<br>Correlation with <i>seconds</i> | <b>AVNN</b><br><b>SDNN</b><br><b>rMSSD</b><br><b>pNN50</b> | Time domain                                                            |
| <b>marker</b>                             | Irrelevant                                                 | <b>TP</b>                                                  | Segment from the end of the T wave to the beginning of the next P wave |

The two variables provided by the sensor GSR for labeling the data, namely, *footGSR* and *handGSR*, are directly eliminated because this information cannot be obtained with the devices used in the proposed system, where only the clock signals are collected. The variables EMG and ECG are also not used because of the difficulty in acquiring them through a portable device, and the attribute *marker* is not used because it is irrelevant to our case study.

Regarding the characteristics that can be extracted from the heart rate (HR) in the frequency domain, to determine the ultra-low frequency (ULF), the recording time has to be long-term, approximately 24 h. [28] Therefore, this type of signal is not usually used in practice and is also discarded for this project because most of the values they present are null, probably because of the short time intervals of the experiment. For data with intervals of very low frequency (VLF), low frequency (LF), and high frequency (HF), the recording time was in the short term, from 1 to 5 min, and in the long-term. However, in practice, the same happens with the information of the VLF band as with the ULF band, in that practically all values take a null value; therefore, this characteristic is discarded for the

development of the model. These results are similar to those of other authors, who have shown that the VLF band is a very unreliable measure for readings of less than 5 min.

A review of the *outliers* and missing values in each attribute shows that there is one variable that has 99% of the data empty; therefore, the variable  $LF - HF$ , which is the low-to-high-frequency power ratio, is removed. Of the remaining attributes, only the attribute *AVNN* has missing values, with the value of a total of 122 instances (3% of the total instances) being unknown. Because this is a minimum percentage, a function is performed to replace these values with the average, as this is a numeric attribute.

Therefore, the set of features for the construction of the model are the HR, respiration (*RESP*), time between wave intervals (*s*, *newtime* y *TP*), time domain (*AVNN*, *SDNN*, *rMSSD*, and *PNN50*), and frequency domain (*LF* and *HF*) features of HR. The N-N intervals take the mean values of HR in a 30 s window or time interval. All of these features are required for feature extraction from HRV. The HRV value is the N-N interval calculated based on the peaks of the waveform, ECG wave peaks. The indicators obtained after analysis in the time and frequency domains reflect physiological characteristics and stress-related information ECG [70].

Time-domain analysis is used to calculate IBI and generate various indicator values, including the mean value, standard deviation, ratio, or differential, which can be used to evaluate the stress. Frequency domain analysis was used to calculate the power spectral density of IBI in order to estimate the power distribution within the frequency range of the overall signal [70].

After the initial processing of the data, the correlations between the variables were studied to determine whether there were any highly correlated attributes that could be eliminated. The variables *interval\_in\_seconds* and *AVNN*, which correspond to the time interval in seconds between two heartbeats and the average of all the intervals between heartbeats, respectively, have a correlation value of 1 obtained using Pearson's correlation coefficient, which measures the linear trend between each pair of numerical variables. This coefficient takes values between 0 and +1; therefore, a value of 1 indicates that the items are highly correlated. In this case, one of the variables is eliminated (*interval\_in\_seconds*) because it does not provide additional information but measures the same characteristic. The same is true for attributes *time* and *seconds*, the first of which is discarded. Despite the fact that there are two other variables with values above 0.7 (*SDNN* and *RMSSD*), they are not removed during the preprocessing stage as this may result in loss of information. The Pearson's feature correlation matrix is shown in Figure 5.



Figure 5. Pearson's correlation matrix of characteristics.

To finalize the preparation of the data for machine learning, a normalization of the input variables to the ML algorithms is performed, in which the values are set within a defined range, and it is avoided that variables measured at different scales contribute differently to the model's fit and learning function, potentially creating a bias. *StandardScaler* from *scikit-learn* was used to transform the data so that the distribution of each attribute ( $\chi$ ) has a mean value of  $\mu = 0$  and a standard deviation of  $\sigma = 1$ , where each input is normalized within defined limits. The mathematical formula for the standardization procedure is given in Equation (6).

$$\zeta = \frac{\chi - \mu}{\sigma} \quad (6)$$

where the mean and standard deviation were calculated using the formulae shown in Equations (7) and (8), respectively.

$$\mu = \frac{1}{N} \sum_{i=1}^n x_i \quad (7)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - \mu)^2} \quad (8)$$

The class considered was the stress level of the user. The interest in the problem is to determine whether the user is in a state of stress; therefore, it is a binary classification model, where the label can take two values: 1 = *stress* or 0 = *no stress*. It was also observed that there was no imbalance between the classes, as there were 2170 samples labeled with stress and 1959 without stress; therefore, the difference in instances between the two classes, approximately 10%, should not affect the classification. Even so, in some algorithms, such as trees, and in some parameters (such as *class\_weight*), a higher minimum value is given to the minority class to compensate for the difference in values.

#### 4.2.2. Feature Extraction

The detection of emotions requires the proper extraction of signal features, which are correlated with the emotional states recorded by the participants in the self-assessment; that is, the relationship between the features and emotions determines the physiological reaction and is used as input for the classifier [69].

In the signals of ECG, which are similar to those of PPG, the time-domain and frequency-domain characteristics are extracted to determine HRV. Parametric measurements of signals ECG in the time domain quantify the variability of IBI successive [69] and are important for the analysis of short-term recordings [28]. In this method, HR is taken at any instant or between intervals, determining either NN or instant HR. The frequency domain determines the power distribution because it distinguishes HR signals according to their frequency and intensity. It provides information on the extent to which the value of HR changes by exploiting periodic oscillations of HR at various frequencies. The main calculated spectral components are denoted by ULF, VLF, LF, and HF. Table 6 lists the features generated from the ECG signal peaks.

#### 4.2.3. Feature Selection

Different feature selection methods can be used to analyze the relevance of a feature and select a subset of attributes. The main goal of feature extraction and selection is to combine the features that best represent the dataset and determine the most important ones.

A classifier-independent feature-filtering method was used to sort the features and assign a score to each feature to estimate the class. The *SelectKBest* class from the *Sklearn* library was used for feature scoring, which removed all but the  $K$  attributes with the highest score, which in practice was set to a value of  $K = 5$ . The result of the best and theoretically most contributing features in the class ranking returns a list of the best attributes: [HR, NNRR, AVNN, AVNN, SDNN, SDNN, RMSSD].

**Table 6.** Time and frequency domain measurements of the signal.

| Time Domain |                                                                                      | Frequency Domain |                                                                                         |
|-------------|--------------------------------------------------------------------------------------|------------------|-----------------------------------------------------------------------------------------|
| AVNN        | Average of all intervals NN.                                                         | ULF              | Ultra-low frequency: total spectral power of all NN intervals down to 0.003 Hz.         |
| SDNN        | Standard deviation of all intervals NN.                                              | VLF              | Very low frequency: total spectral power of all NN intervals between 0.003 and 0.04 Hz. |
| rMSSD       | Root mean square differences between adjacent NN intervals.                          | LF               | Low frequency: total spectral power of all NN ranges between 0.04 and 0.15 Hz.          |
| pNN50       | Percentage of differences between adjacent NN intervals that are greater than 50 ms. | HF               | High frequency: total spectral power of all NN intervals between 0.15 and 0.4 Hz.       |

#### 4.2.4. Dataset Division

Once the data have been preprocessed and feature extraction and selection have been performed, the dataset is divided into two subsets, one for training and the other for testing, for the correct development and learning of the model ML.

Therefore, the dataset is split to make the model as generalizable as possible because having the training and test sets allows the model to be evaluated on unknown data or data that have not been used for learning.

For data splitting, Sklearn's *train\_test\_split* function is used, which splits the data into random subsets of training (*train*) and testing (*test*), where 20% of the data are used for testing and the remaining 80% for training.

#### 4.3. Training and Validation of the Model

Different supervised algorithms were implemented, and some of them were tested with different parameters to find the model that best performed the classification. Regardless of the algorithm used, cross-validation was performed during training to check whether the model was valid by introducing inputs from a new dataset and avoiding overfitting.

##### 4.3.1. Gaussian classifier

The first model developed implemented the NB algorithm based on probabilities and used for classification. It uses the variables that store the training and test data in Section 4.2.4 because the class must be separated. An instance of the *GaussianNB()* classifier was created using the Sklearn library, and the model was trained by passing the values of the list of selected features obtained in Section 4.2.3 and the labels or classes corresponding to each instance. Once the model has been trained, the classification accuracy is tested with the training data, and the predictions made by the model are compared with the training data with real labels. An accuracy of 56% was obtained for the training set. The accuracy of the model using the test dataset was 55%. The results of the different evaluation metrics indicate that the model does not classify correctly, as an accuracy value of 56% is obtained, with a recall of 55%, F1 Score of 56%, and ROC of 56%.

##### 4.3.2. KNN

The KNN or K-nearest neighbors algorithm is based on instances and is used to classify values by searching for the most similar data by closeness learned during training. For this reason, it is very important to select the value or number of the nearest elements. For this reason, a transformation of the features was previously carried out, scaling them within a range and looking for the value to give to parameter *k*. For this purpose, we compared the error rate with different values of *k*, choosing a final value of 7, which is the one that achieves the best accuracy. Before training the model, preprocessing is carried out for the application of the algorithm, which consists of transforming the characteristics by scaling each one of them within a range so that all the variables have the same weight.

Once the model is created, it is trained and the classification result is validated. An accuracy of 80% was obtained for the training set and 71% for the test data. The confusion matrix returns the values shown in Table 7. The value returned by *Recall* is 71%, as is the *F1*

*Score*. The classifier was tested using the data collected by the sensor, and an accuracy of 87% was achieved.

**Table 7.** KNN confusion matrix.

|      |           | FORECAST  |           |
|------|-----------|-----------|-----------|
|      |           | Positives | Negatives |
| FACT | Positives | 327       | 141       |
|      | Negatives | 156       | 409       |

#### 4.3.3. Random Forest

Random Forest is based on a large number of individual decision trees that operate as a set. Each tree returns a prediction of the class within the forest, and the prediction of the model corresponds to the class with the highest votes. The model is created with *RandomForestClassifier*, a number of 100 trees is set as a parameter, and *sqrt* is used as a method for selecting the maximum number of features for each tree, of which the number to be considered in each split is performed automatically. Samples were used for the construction of the trees, the model was trained, and the classification results were observed. The confusion matrix shows that the model classifies 459 instances correctly of stress as stress and 104 incorrectly as non-stress being stress. On the other hand, the model classifies 591 instances correctly as non-stress, whereas it classifies 156 instances as stress when they are not, as shown in Table 8. The model's classification accuracy is 84%, *Recall* returns a value of 82%, and *F1 Score* is 83%.

**Table 8.** RF confusion matrix.

|      |           | FORECAST  |           |
|------|-----------|-----------|-----------|
|      |           | Positives | Negatives |
| FACT | Positives | 459       | 104       |
|      | Negatives | 156       | 591       |

#### 4.3.4. Bagging

The bagging algorithm fits classifiers on random subsets of the original dataset and then aggregates their individual predictions. This is used to reduce the variance in the decision tree. Training was performed on the data using *10-fold cross-validation* to ensure that the results were independent of the partition between the training and test data. An accuracy of 87.47% was obtained as it classifies 3612 instances correctly and 517 instances incorrectly. *Recall* and *F1 Score* return a value of 87.5%, while *ROC* returns 94.4%. Looking at the confusion matrix, the model classifies 1631 instances correctly as stress and 337 incorrectly as non-stress being stress, as shown in Table 9. However, the model classifies 1981 instances correctly as non-stress being non-stress, while classifying 180 instances as stress when they are not.

**Table 9.** Bagging confusion matrix.

|      |           | FORECAST  |           |
|------|-----------|-----------|-----------|
|      |           | Positives | Negatives |
| FACT | Positives | 1631      | 337       |
|      | Negatives | 180       | 1981      |

#### 4.3.5. Multi-Layer Perceptron

A neural network is created based on a four-layer hidden perceptron. After instantiating and training the model, the weights of the individual neurons were adjusted to learn

from the labeled dataset, as shown in Table 10. Several settings of the learning algorithm were tested, and the final model achieved an accuracy of 78.56%. *Recall* and *F1 Score* return a value of 78.4% and ROC is 85%. It classifies 1407 instances as stress correctly and 561 as stress without being stressed; on the other hand, it classifies 1837 instances as non-stress correctly and 324 instances as stress that are not really stress.

**Table 10.** MP confusion matrix.

|      |           | FORECAST  |           |
|------|-----------|-----------|-----------|
|      |           | Positives | Negatives |
| FACT | Positives | 1407      | 561       |
|      | Negatives | 324       | 1837      |

#### 4.3.6. Decision Trees

Decision trees are graphical representations of possible solutions to a decision based on certain conditions and are one of the most widely used algorithms for classification. For DT, the entropy was set as a parameter to measure the number of splits for the maximum depth of the tree. The *KFold* function was used to determine the accuracy as a function of the value of each variable. This function creates several subgroups from the initial input data to evaluate and validate the trees with different depth levels to choose the best result. In our case, a maximum depth of 20 was chosen, and the remaining parameters were set to the values listed in Table 11. *class\_weight* = 1 : 1.2 indicates that weight 1 is selected for the stress class and weight 1.2 for the non-stress variable in order to compensate for the number of extra instances of the first class. Some of these parameters were tested to determine which values were optimal for the model and classification. The accuracy achieved with DT was 90.34%, and the validation metrics indicated that the model was classified correctly, obtaining an accuracy value of 90.34%, *Recall* 88%, *F1 Score* 86%, and ROC 89%.

#### 4.3.7. SVM

Finally, a support vector machine model was tested; however, it obtained a very low classification accuracy of 63.98%. *Recall* returned a value of 63.9%, and *F1 Score* and ROC returned a value of 0.637%. The confusion matrix shows that the model correctly classifies 1126 instances as stress being stress and 842 incorrectly as non-stress being stress. On the other hand, the model correctly classified 1512 instances as non-stress, whereas it classified 649 instances as stress when they were not.

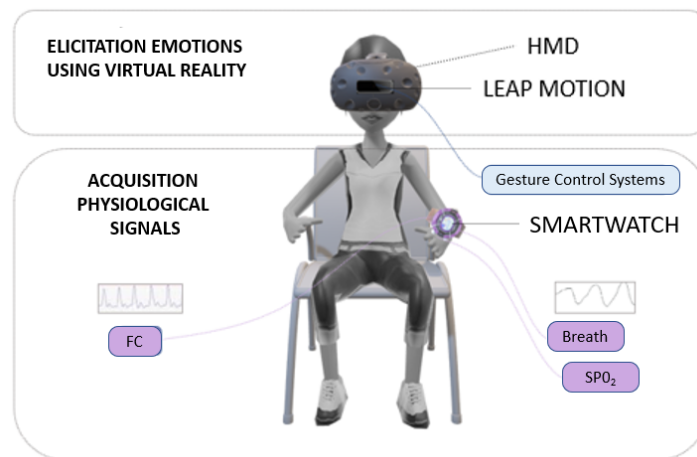
To summarize this section, different classification models have been presented, which were trained using the dataset described in Section 4.1 and the algorithms and parameters listed in Table 11. These trained models were validated through a case study in immersive scenarios developed in this work, as described in the following section.

**Table 11.** Parameters set in the algorithms.

| Algorithm | Parameter     | Value        | Algorithm | Parameter    | Value |
|-----------|---------------|--------------|-----------|--------------|-------|
| KNN       | Normalization | MixMaxScaler | NB        | K            | 7     |
|           | Neighbours    | 7            |           |              |       |
| DT        | Splits        | 10           | RF        | Splits       | 2     |
|           | Leaf          | 6            |           | Leaf         | 1     |
|           | max_depth     | 20           |           | max_features | sqrt  |
|           | class_weight  | {1:1.2}      |           | estimators   | 100   |
|           | Criteria      | entropy      |           | Criteria     | GINI  |

## 5. Case Study

Once a classification model is trained, its goal is to be tested in a real-world scenario for stress detection and model validation. For this purpose, an immersive system based on VR was developed. Emotions were induced by different stimuli provoked during a simulation to obtain the physiological response data with the clock, as shown in Figure 6. People are monitored while they perform a simulation of natural driving and two situations designed with the intention of generating stress, such as a head-on collision with a car and a vehicle rollover. A Garmin Forerunner 235 watch was used to collect the physiological responses of users. This information will result in a new dataset that will be applied to the model, as it tests the effectiveness of the classifier that, based on the user's status, decides whether the simulated situation has caused stress.



**Figure 6.** Proposed system structure.

The models resulting from the application of different algorithms were used to predict and label the stressed or unstressed states in the new unlabeled dataset. The results obtained by the neural network were compared with the user responses at the end of the test regarding the stressful stages experienced.

### 5.1. Signal Capture Device

To capture physiological signals and, in particular, those related to sensor PPG, which is integrated in some watches and obtains information about the HR, the most commonly used devices by other authors were reviewed to obtain the HRV. A Garmin Forerunner 235 was used for user monitoring. The PPG signals were recorded using the optical sensor of the watch at a sampling rate between 1 and 2 s. The clock consists of three green LEDs and one infrared LED, as shown in Figure 7, which allows the calculation of the HR through volumetric variations in the blood circulation.



**Figure 7.** Garmin signal capture device.

The Leap Motion device was also used (Leap Motion: <https://www.ultraleap.com/product/leap-motion-controller/>, accessed on 8 May 2023) to capture and track in real time the hand movements made by the user. In this way, users can see their hands during the simulation and interact with other objects in the environment, thereby providing greater

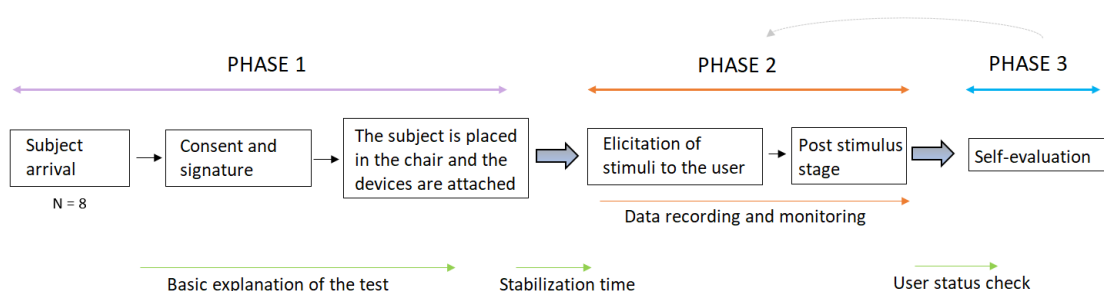
credibility. In addition, this device controls changes in the position and movement of the hands during the entire test. This is key because, in this sense, the accelerometer data are not as important as in other studies because the individual performs the entire experiment seated, so the accuracy of labeling is maintained, as the possible movements of the user are not misinterpreted as physical stress. In contrast, gestures or movements made with the hands can provide interesting information.

### 5.2. Elicitation of Emotions

For the elicitation of stress-determining emotions, a VR application by using the game engine created by Epic Games under the name of Unreal Engine 4 (Unreal Engine 4: <https://www.unrealengine.com/en-US/>, accessed on 8 May 2023), with version 4.24. The software is based on the C++ language and includes advanced features, such as a dynamic and real-time lighting system and a powerful graphics engine for rendering 2D and 3D graphics, which are fundamental to the credibility of the simulation. The physics engine allows the approximate simulation of the physics of objects; the audio engine is responsible for the treatment, modification, and output of sound, simulating effects such as echo indoors or the Doppler effect when the sound source is in motion, and it has algorithms IA, such as the behavior of the non-player character (NPC), whose movements are controlled by computer algorithms. In addition, it is compatible with numerous platforms and is a powerful engine for VR. The HMD HTC Vive Pro (HTC Vive: <https://www.vive.com/mx/product/vive-pro/>, accessed on 8 May 2023) was used.

The developed application consists of a driving and traffic accident simulator in which the user is the co-driver of the vehicle. In this way, emotions are induced by different stimuli provoked during the simulation with the objective of obtaining physiological response data with the watch. The collected information is used to test the effectiveness of the classifier.

The development of the application schematically encompasses several processes to stimulate the user: the design and modeling of the three-dimensional environment and animation, the programming of events in the simulation, and the configuration of the environmental context. Emotion recognition focused on participants' reactions to the prepared stimuli and the environment developed in the simulation. In this study, we developed an application for the simulation of an accident and the acquisition of relevant sensations to evaluate the effect of the simulated content on users. Figure 8 shows the experimental phases.



**Figure 8.** Design and experimental protocol used.

### 5.3. Scenarios

The developed software consists of a main menu covering a base scenario, shown at Figure 9, and other two different three-dimensional scenarios, the simulation of a head-on collision and a rollover with a vehicle. In both cases, the user was positioned in a chair and fitted with the VR headset and goggles. During the execution of the program, sounds relevant to different actions were emitted.

The user starts the game in the co-driver's seat inside the car, together with the driver, who will start explaining to the user the test he is going to face and the required safety actions. During this brief introduction, the co-driver can visualize the entire environment

through the HMD, namely, the interior of the car positioned in a parking lot with more vehicles heading towards the road. Subsequently, the vehicle starts to drive, progressively increasing its speed, and travels normally. The aim is to induce a basic or neutral emotional state. At this point, the development of the game varies depending on the two existing scenarios.



**Figure 9.** Base case scenario.

#### 5.3.1. Crash Simulation

In the case of the vehicle crash simulation, the car is started, gets on the road, and moves forward naturally while the driver talks calmly, and at a certain point, the driver warns and recreates a “brake failure” and warns and shouts that “he cannot turn”. After a few seconds, the car crashes into a billboard or an advertising pole, as shown in Figure 10. The car’s front window shatters, the front airbag deploys, and the driver is propelled forward and, braked by the airbag, suffers an overshoot to the rear. The co-driver’s or participant’s airbag is also deployed, and the crash is reinforced by the relevant sounds. After a few seconds of waiting, the screen goes dark, the user is asked if he/she is OK, and the participant vacates the chair, which can be used by a new user, or the second scenario is executed.



**Figure 10.** Vehicle crash scenario.

#### 5.3.2. Rollover Simulation

When the event is the simulation of the vehicle rollover, the beginning is similar and normal driving occurs, in which the driver communicates information about the landscape to the user, and, at a specific moment, the user loses sight of the road by looking at the co-driver while talking and this causes the driver to get on a ramp in which he will go a distance in a straight line, and, after a few seconds, the vehicle begins to lean laterally and to wobble and finally ends up overturning, as shown in Figure 11. The impact with the ground causes the glass to fracture and the appearance of the co-driver’s side and front airbags. As in the previous case, after a few seconds, the screen goes black, and the user is asked about their status. The participant is either off the platform, leaving it free for another user, or may experience a vehicle crash.



**Figure 11.** Vehicle rollover scenario.

#### 5.3.3. Simulation Recording with Users

For each of the eight participants, a session of 90 s each was recorded, resulting in a total of eight sessions. Although the initial idea was to test each participant in both scenarios, some users did not want to try the second scenario after completing the first scenario. In other cases, the results collected by the sensors depended on which scenario was performed first because the stimuli were very similar. Therefore, it was decided that half of the users would perform the first experiment, and the other half would perform the second. Two sessions had to be discarded and repeated because the signals were captured as soon as the HMD was placed on the user without allowing a few seconds for the user to calm down, and only two users had previously used the HMD. Finally, in the first few minutes of recording, a blank screen was shown with a musical stimulus intended to induce a neutral emotional state, and once the minute had elapsed, users were immersed in the proposed scenarios. The data acquisition protocol was similar for each participant.

On the other hand, it is difficult to evaluate the scenarios and label the data, so we had to resort to a questionnaire to determine the state of stress of the users during the collection of signals. To this end, each user reported and filled out a questionnaire with the parts of the simulation that had caused them stress, allowing the user to locate approximately the periods of time when the user felt stressed, since the perception of the emotional state is subjective or dependent on the user, and the HRV varies depending on the test and tolerance of each person.

#### 5.3.4. Application Method

Despite the previous study of the participants who were going to participate in the experiment, before starting the recording of each participant, they were asked to fill in a questionnaire that collected general information about their state of health. In order to avoid alterations in the datasets, users had to affirm that they had not consumed alcoholic beverages or coffee, that they had not exercised before the test, and that they had no previous illnesses or pathologies. In addition, the users signed the consent to perform the test and the risks that the development of the test may cause, such as possible dizziness, anxiety attacks, or other types of possible factors caused by the use of the HMD and even more so by the high character of the test, because the simulation of an accident is a highly startling situation. The participants were then provided with a set of instructions explaining the experimental protocol without revealing what happened in the simulation, and the driver interacted and informed the user during the simulation.

Once the safety instructions and information necessary to perform the test have been provided, the user is seated in a chair, the wrist band is placed on the wrist, the HMD is placed at eye level on the head, together with the sound headphones, and the simulation is started in the different scenarios. To avoid distractions and increase concentration, only the user and the person supervising the test were in the room.

#### 5.3.5. Participants

For the selection of participants, a study of the profiles of the subjects suitable for the experiment was conducted as it has been found that there are significant differences in the experience of emotions in different individuals [71]. It was revealed that the age, gender,

personality, or health of the participants, especially those undergoing treatment or regularly taking medication, may affect the test results [25]. In our case, we excluded profiles of subjects with cardiac diseases or problems, as they could interfere with physiological signals, and people with vertigo or dizziness, which may be enhanced when carrying out the VR simulation. Eight healthy people (four men and four women), ranging in age from 24 to 50 years, participated in this experiment voluntarily and signed a consent form.

#### 5.4. Display of Results

The visualization of the results is performed using a mobile application developed for Android, whose interface is shown in Figure 12. The software contains the stress recognizer system, which enters the characteristics as input to the system by accessing the information provided by the watch sensor, making a connection between the data collected from the sensors and the mobile application, which returns the user's status as stressed or unstressed.

The application records, traces, exports, and tells the emotional state of the HRV characteristics in the time and frequency domains and extracts, plots, stores, and exports heart rate variability functions in the heart rate, rr, time, and frequency domains (AVNN, SDNN, rMSSD, pNN50, LF, HF, LF/HF).



Figure 12. Mobile application interface.

#### 6. Results

The results after applying the different algorithms are shown in Table 12. It shows the classification accuracy of the algorithm and the value returned by the different evaluation metrics.

The KNN algorithm was chosen as the mobile application classifier because it is one of the simplest classification algorithms, and despite not being the best performer, the results are highly competitive, as shown in Table 12, with a classification accuracy of 87.02%. The overall performance of the classifier, summarized over all possible thresholds, is given by the ROC curve. This metric obtains a value of 87%, which means that the classifier is effective at separating the instances of the two classes and identifying the threshold that best separates them.

The choice of the KNN algorithm was based on the type of problem to be solved and the dataset used. As the number of model inputs is not very high because only the PPG sensor is available for data capture, it can be considered an advantage that the KNN algorithm is based on instances. Thus, when making the prediction, the algorithm is based on the instances trained to perform the classification using these data to generate the

response. Therefore, although it does not build any model, as in the case of DT, it classifies when a test instance arrives without assuming the distribution of the data. In addition, the low dimensionality or small number of attributes taken as input to the algorithm and the feature selection performed as preprocessing of the data prevents the accuracy from being affected by irrelevant features or noise.

However, although the computational cost of this algorithm is high because it stores all the training data and requires a lot of processing resources (CPU) and a large amount of memory, it has few input features and deals with short-time recordings. Therefore, it was not a problem in this case study, nor would it be a problem for implementation in a mobile device, where no delay is found in the classification. Additionally, since no storage problems are encountered because large amounts of data are not handled, it is not necessary to carry out any feature reduction procedures (such as PCA). For these reasons, the KNN algorithm was chosen as the classifier based on the understanding that the best model of the data is the data itself without seeking an optimized model. Rather, each instance was compared with the training data to obtain a criterion and measure the similarity of each instance to be classified.

The DT algorithm was not chosen for the classifier despite having the highest accuracy (90.34%) because in many cases it is unstable in classification. Small variations in the data can cause large changes in the tree structure. In the stress detection during the simulation, the data change had a wide range of values, which can lead to oscillations in the classification.

However, RF and *bagging* usually improve the results of DT, and this is not the case for the results obtained, as shown in Table 12. This can be interpreted as overfitting of the model; that is, the DT algorithm may not generalize well from the training data, and the prediction accuracy with the sensor data may not return accurate results. In addition, one of the great advantages of DT is its ability to identify important variables in high dimensionality problems or the large number of values that the target variable can take, which is not the case for the dataset used; therefore, they are not relevant in our study. Furthermore, decision trees make locally optimal decisions at each node but do not guarantee that the global tree returned is the most optimal.

The bagging algorithm is often used to reduce the variance in the DT because the value of the variance can mean that, by randomly dividing the training data into two groups, if a DT is fitted to each half, the results obtained can be quite different, obtaining an accuracy very similar to the KNN algorithm.

**Table 12.** Ranking results with different algorithms.

| Algorithm             | Precision | Recall | F1 Score | ROC  |
|-----------------------|-----------|--------|----------|------|
| Decision Tree         | 90.34%    | 0.88   | 0.86     | 0.89 |
| Bagging               | 87.47%    | 0.87   | 0.87     | 0.94 |
| Random Forest         | 84.13%    | 0.82   | 0.85     | 0.83 |
| K-Nearest Neighbor    | 87.02%    | 0.71   | 0.71     | 0.87 |
| Multilayer Perceptron | 78.56%    | 0.78   | 0.78     | 0.85 |
| SVM                   | 63.98%    | 0.63   | 0.63     | 0.63 |
| Gaussian Näive Bayes  | 56.00%    | 0.55   | 0.56     | 0.56 |

After developing the model, tests were conducted on the data collected using the watch. The model receives the information acquired from the physiological signals as inputs and returns the result of the classification via the mobile device, indicating whether the subject is under stress.

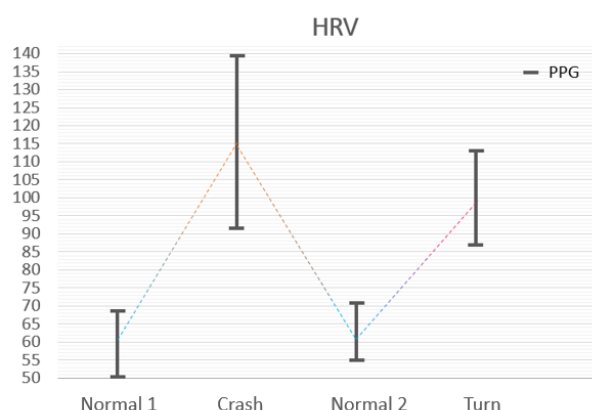
The result of the model was checked with the self-assessment made by the user, which indicated the stages he/she felt stressed and the stages he/she was in a neutral state. All subjects reported feeling stressed during the most notable stimulus, the accident; in fact, Table 13 shows the mean values of the resultant HR obtained using the sensor PPG across all participants at each experimental stage, and Figure 13 shows the HRV over time as a

function of each of the stimulus presentation stages. These results clearly show that the stimulus elicited a physiological response in participants.

**Table 13.** Average heart rate obtained in each experimental step.

| Event    | PPG     |
|----------|---------|
| Normal 1 | 53 bpm  |
| Crash    | 103 bpm |
| Normal 2 | 55 bpm  |
| Turn     | 97 bpm  |

However, other scenarios experienced during the simulation also caused stress and were correctly detected by the model. Some participants reported the first few seconds as stressful, while the driver of the vehicle explained the test and when the car increased its speed significantly. Others reported that the driver's raised tone of voice and the moments before the accident caused stress. The model returns a time series with the result of the user's emotional state and is checked against the stages reported by the participants. The model is able to mark the stages reported by the user as stress and to detect the non-stress class at times when the user is in a neutral emotional state.



**Figure 13.** Variation of HR during stimulus presentation stages.

## 7. Conclusions and Future Work

After investigating methods for measuring human emotional states, it was found that with current consumer technology, simply by capturing information about HRV and applying machine learning techniques, it is possible to develop a system capable of reliably detecting stress, as shown in the results obtained in Section 6.

The system proposed in this paper is a machine learning model capable of determining whether users are under stress with high accuracy and in real time using physiological signal data based on the HRV obtained from the PPG signals captured by commonly used low-cost portable watches. Different algorithms were implemented, and it was shown that HRV is valid for classifying user stress. An accuracy of 90.34% was obtained for stress detection using DT for eight participants.

The results show that HR correlates closely with the stress level of virtual vehicle occupants. Physiological signals captured through a commonly used watch can provide a metric of driver stress and the ability to monitor people in cars and can gather useful information on how different road conditions affect drivers.

It should be noted that performing the experiment in a seated position minimizes the risk of PPG sensor failure owing to external conditions, such as irregular movements, but in other working conditions, the results may be altered. It should also be noted that many attribute values were null because the experiment was conducted over a short time interval.

In addition, HRV can vary between subjects as it is not comparable for every individual but depends on factors such as age, gender, health status, and consumption, and that each person is subjective or responds differently to similar stimuli. Even so, HRV can become a valuable noninvasive method for the daily assessment of people's health status.

Although NB usually gives good results in classification, in this case, it was the algorithm that performs the worst in class classification. The same occurs with SVM, which, despite the good results obtained in other studies with similar objectives, achieved an accuracy of 63.98% in our case. By contrast, both RF and DT classified stress classes with an accuracy of over 84%. The best result was obtained with DT, with an accuracy of 90.34%, indicating that the system was able to detect stress, even when testing the model with data collected through the PPG sensor of a watch during the simulation. The dataset proved to be valid for classifying the user's emotion based on the data collected by the sensors despite the fact that few input features were handled in the work carried out. The results also show that stimuli induced by VR technology elicit human physiological responses.

Table 14 presents a comparison of this study with other studies with similar objectives. The high accuracy of the classification can be observed, allowing daily monitoring with freedom of movement at low cost and with immersive stimuli that can be changed at any time without requiring complex installation.

**Table 14.** Project Comparison with other works (V: True, X: False).

|                 | Real-Time Monitoring | Low Cost | Immersive Stimulation | Freedom of Movement | Easy Installation | Lab Experiment | Accuracy >85% |
|-----------------|----------------------|----------|-----------------------|---------------------|-------------------|----------------|---------------|
| <b>Proposal</b> | V                    | V        | V                     | V                   | V                 | V              | V             |
| [53]            | V                    | V        | X                     | V                   | V                 | V              | X             |
| [22]            | V                    | V        | X                     | X                   | V                 | X              | V             |
| [23]            | X                    | X        | X                     | X                   | X                 | X              | V             |
| [20]            | V                    | V        | X                     | V                   | V                 | V              | X             |
| [13]            | X                    | V        | V                     | X                   | X                 | X              | X             |

This work is extensive and involves multiple studies on the design and development of machine learning modules, as detailed in this article. It is a preliminary investigation; therefore, the results and discussion of the classifier system are very limited as they have been performed on a small number of individuals, and only one emotion was analyzed in two extreme situations. The tasks undertaken in the course of this work have led to the achievement of several main objectives, which can be summarized as follows:

- A comprehensive review of the state of the art is carried out.
- A review and analysis of the different classification methods based on ML techniques and model evaluation metrics are carried out.
- Several models are trained on the dataset generated through the driving simulations.
- VR software is developed to induce stress states in the user.
- A commercial smartwatch is used to capture and acquire physiological signals.
- An application is made to visualize the results on a mobile device.
- The experiment to validate the classifier is designed and carried out.
- An analysis of the results obtained is carried out.

In future work, data can be collected from more non-invasive sensors to record more information and detect more emotional states, which can be very beneficial in many areas of daily life. Many of the situations that are tested, such as accidents or dangerous situations, can only be tested using this simulation tool. This turns out to be one of the strengths of the developed tool. In the future, the prototype will improve aspects of user interaction as well as its evaluation in a real environment.

Following this line of work, the induction of stimuli could be enhanced by changing the office chair to a mechanical platform that accompanies and gives more realism to the actions carried out by the player in the VR application. A connection is made between the

VR application and the programmable automaton, and the physical sensations that may be associated with the environment can be reproduced.

It also considers the development of a system that directly acquires the user's signals while performing everyday actions; thus, the raw features can be input to the emotion recognition model; that is, they directly become part of the training dataset, and the results can be displayed directly on the mobile device. In this way, emotions can be analyzed in multiple situations. However, it would also be interesting to experiment with a larger number of individuals so that the results of the classifier system are not very limited.

The main contribution of this work is that this study shows the possibility of including recommendations to users based on their moods in more ambitious projects, as there is a correlation between certain affective states and certain places or people, as has been demonstrated. Although it should be considered that the system is developed to run on a mobile system, the model should not incur excessive computational cost.

**Author Contributions:** Conceptualization, A.-B.G.-G., A.L.-R. and B.P.-L.; methodology, N.M.-G., A.-B.G.-G., A.L.-R. and B.P.-L.; software, N.M.-G.; writing—original draft preparation, N.M.-G., A.-B.G.-G., A.L.-R. and B.P.-L.; writing—review and editing, A.-B.G.-G., A.L.-R. and B.P.-L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the project RTI2018-095390-B-C32 (MCIU/AEI/FEDER, UE) and by the project “COordinated intelligent Services for Adaptive Smart areaS (COSASS), Reference: PID2021-123673OB-C33, financed by MCIN/AEI/10.13039/501100011033/FEDER, UE.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|     |                              |
|-----|------------------------------|
| 3D  | 3 Dimensional                |
| ACC | Accelerometer                |
| AI  | Artificial Intelligence      |
| ANN | Artificial Neural Networks   |
| AUC | Area under the ROC Curve     |
| BPV | Blood Pressure Variability   |
| BVP | Blood Volume Pulse           |
| CAS | Context Aware System         |
| CNN | Convolutional Neural Network |
| CSV | Comma Separated Value(s)     |
| DNN | Dynamic Neural Network       |
| DT  | Decision Tree                |
| ECG | Electrocardiogram            |
| EDA | Electrodermal Activity       |
| EEG | Electroencephalogram         |
| EMG | Electromyography             |
| FN  | False Negative               |
| FP  | False Positive               |
| FR  | Facial Recognition           |
| GPS | Global Positioning System    |
| GSR | Galvanic Skin Response       |
| HF  | High Frequency               |
| HMD | Head-Mounted Display         |
| HNB | Hidden Naïve Bayes           |
| HR  | Heart Rate                   |
| HRV | Heart Rate Variability       |
| HT  | Hoeffding Tree               |
| IBI | Interbeat Interval           |

|      |                                         |
|------|-----------------------------------------|
| IoT  | Internet of Things                      |
| KNN  | K-Nearest Neighbor                      |
| LF   | Low Frequency                           |
| ML   | Machine Learning                        |
| MP   | Multilayer Perceptron                   |
| NB   | Naïve Bayes                             |
| NPC  | Non-Player Character                    |
| OS   | Oxygen Saturation                       |
| PCG  | Phonocardiogram                         |
| PPG  | Photoplasmography                       |
| RF   | Random Forest                           |
| ROC  | Receiver Operating Characteristic Curve |
| RP   | Respiratory Pattern                     |
| SC   | Skin Conductivity                       |
| SLR  | Systematic Literature Review            |
| SPO2 | Pulse Oximetry                          |
| SR   | Speech Recognition                      |
| SVM  | Support Vector Machines                 |
| TN   | True Negative                           |
| TP   | True Positive                           |
| ULF  | Ultra-Low Frequency                     |
| VLF  | Very-Low Frequency                      |
| VR   | Virtual Reality                         |

## References

1. Zhang, J.; Yin, Z.; Chen, P.; Nichele, S. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Inf. Fusion* **2020**, *59*, 103–126. [\[CrossRef\]](#)
2. Gruenewald, A.; Kroenert, D.; Poehler, J.; Brueck, R.; Li, F.; Littau, J.; Schnieber, K.; Piet, A.; Grzegorzec, M.; Kampling, H.; et al. Biomedical Data Acquisition and Processing to Recognize Emotions for Affective Learning. In Proceedings of the 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, Taiwan, 29–31 October 2018; pp. 126–132. [\[CrossRef\]](#)
3. Yang, H.; Han, J.; Min, K. A Multi-Column CNN Model for Emotion Recognition from EEG Signals. *Sensors* **2019**, *19*, 4736. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Hsu, J.L.; Zhen, Y.L.; Lin, T.C.; Chiu, Y.S. Affective content analysis of music emotion through EEG. *Multimed. Syst.* **2017**, *24*, 1–16. [\[CrossRef\]](#)
5. Seo, J.; Laine, T.; Sohn, K.A. An Exploration of Machine Learning Methods for Robust Boredom Classification Using EEG and GSR Data. *Sensors* **2019**, *19*, 4561. [\[CrossRef\]](#)
6. World Health Organization. *Global Status Report on Road Safety 2018*; World Health Organization: Geneva, Switzerland, 2018; p. 403.
7. Brookhuis, K.A.; De Waard, D. Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* **2010**, *42*, 898–903. [\[CrossRef\]](#)
8. Gao, Z.; Li, C.; Hu, H.; Zhao, H.; Chen, C.; Yu, H. Experimal study of young male drivers' responses to vehicle collision using EMG of lower extremity. *Bio-Med. Mater. Eng.* **2015**, *26*, S563–S573. [\[CrossRef\]](#)
9. Haouij, N.; Poggi, J.M.; Ghalila, S.; Ghozi, R.; Mériem, J. AffectiveROAD System and Database to Assess Driver's Attention. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, Pau, France, 9–13 April 2018.
10. Kajiwar, S. Evaluation of driver's mental workload by facial temperature and electrodermal activity under simulated driving conditions. *Int. J. Automot. Technol.* **2014**, *15*, 65–70. [\[CrossRef\]](#)
11. World Health Organization. *Doing What Matters in Times of Stress: An Illustrated Guide*; World Health Organization: Geneva, Switzerland, 2020; p. 126.
12. Ali, M.; Mosa, A.; Al Machot, F.; Kyamakya, K. Emotion Recognition Involving Physiological and Speech Signals: A Comprehensive Review. In *Recent Advances in Nonlinear Dynamics and Synchronization: With Selected Applications in Electrical Engineering, Neurocomputing, and Transportation*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 287–302. [\[CrossRef\]](#)
13. Yan, L.; Wan, P.; Zhu, D. The Induction and Detection Method of Angry Driving: Evidences from EEG and Physiological Signals. *Discret. Dyn. Nat. Soc.* **2018**, *2018*, 3702795. [\[CrossRef\]](#)
14. Zero, E.; Bersani, C.; Zero, L.; Sacile, R. Towards real-time monitoring of fear in driving sessions. *IFAC-PapersOnLine* **2019**, *52*, 299–304. [\[CrossRef\]](#)
15. Birjandtalab, J.; Cogan, D.; Pouyan, M.B.; Nourani, M. A Non EEG Biosignals Dataset for Assessment and Visualization of Neurological Status. In Proceedings of the 2016 IEEE International Workshop on Signal Processing Systems (SiPS), Dallas, TX, USA, 26–28 October 2016; pp. 110–114.

16. Pinto, J.F.; Fred, A.; Plácido da Silva, H. Biosignal-Based Multimodal Emotion Recognition in a Valence-Arousal Affective Framework Applied to Immersive Video Visualization. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; Volume 2019. [\[CrossRef\]](#)
17. Mavridou, I.; McGhee, J.T.; Hamed, M.; Fatoorechi, M.; Cleal, A.; Ballaguer-Balester, E.; Seiss, E.; Cox, G.; Nduka, C. FACETEQ interface demo for emotion expression in VR. In Proceedings of the 2017 IEEE Virtual Reality (VR), Los Angeles, CA, USA, 18–22 March 2017; pp. 441–442. [\[CrossRef\]](#)
18. Reichenberger, J.; Pfaller, M.; Forster, D.; Gerczuk, J.; Shiban, Y.; Mühlberger, A. Men Scare Me More: Gender Differences in Social Fear Conditioning in Virtual Reality. *Front. Psychol.* **2019**, *10*, 1617. [\[CrossRef\]](#)
19. Eudave, L.; Valencia, M. Physiological response while driving in an immersive virtual environment. In Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Eindhoven, The Netherlands, 9–12 May 2017; pp. 145–148.
20. Zhao, B.; Wang, Z.; Yu, Z.; Guo, B. EmotionSense: Emotion Recognition Based on Wearable Wristband. In Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, China, 8–12 October 2018; pp. 346–355.
21. Nam, S.H.; Lee, J.Y.; Kim, J.Y. Biological-Signal-Based User-Interface System for Virtual-Reality Applications for Healthcare. *J. Sens.* **2018**, *2018*, 9054758:1–9054758:10. [\[CrossRef\]](#)
22. Akbulut, F.P.; Ikitimur, B.; Akan, A. Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome. *Artif. Intell. Med.* **2020**, *104*, 101824. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Yang, H.; Han, J.; Min, K. Distinguishing Emotional Responses to Photographs and Artwork Using a Deep Learning-Based Approach. *Sensors* **2019**, *19*, 5533. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Wang, Y.; Wang, T.; Gong, P.; Wu, Y.; Ye, C.; Li, J.; Ma, T. A multi-label learning method for efficient affective detection. In Proceedings of the 2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), Orlando, FL, USA, 16–19 February 2017; pp. 61–64. [\[CrossRef\]](#)
25. Egger, M.; Ley, M.; Hanke, S. Emotion Recognition from Physiological Signal Analysis: A Review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55. [\[CrossRef\]](#)
26. Granato, M.; Gadia, D.; Maggiorini, D.; Ripamonti, L.A. Feature Extraction and Selection for Real-Time Emotion Recognition in Video Games Players. In Proceedings of the 2018 14th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 26–29 November 2018; pp. 717–724. [\[CrossRef\]](#)
27. Rabhi, Y.; Mrabet, M.; Fnaiech, F. A facial expression controlled wheelchair for people with disabilities. *Comput. Methods Programs Biomed.* **2018**, *165*, 89–105. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Montesinos, V.; Dell’Agnola, F.; Arza, A.; Aminifar, A.; Atienza, D. Multi-Modal Acute Stress Recognition Using Off-the-Shelf Wearable Devices. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2196–2201. [\[CrossRef\]](#)
29. Zużewicz, K.; Roman-Liu, D.; Konarska, M.; Bartuzi, P.; Matusiak, K.; Korczak, D.; Lozia, Z.; Guzek, M. Heart rate variability (HRV) and muscular system activity (EMG) in cases of crash threat during simulated driving of a passenger car. *Int. J. Occup. Med. Environ. Health* **2013**, *26*, 710–723. [\[CrossRef\]](#)
30. Althobaiti, T.; Katsigiannis, S.; West, D.; Bronte-Stewart, M.; Ramzan, N. Affect Detection for Human-Horse Interaction. In Proceedings of the 2018 21st Saudi Computer Society National Computer Conference (NCC), Riyadh, Saudi Arabia, 25–26 April 2018; pp. 1–6. [\[CrossRef\]](#)
31. Arcentales V., A.; Raza, M.; Giraldo, B.F. Characterization of HRV and QRS slope during audiovisual stimulation. In Proceedings of the 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Pucon, Chile, 18–20 October 2017; pp. 1–4. [\[CrossRef\]](#)
32. El-Amir, M.M.; Al-Atabany, W.; Eldosoky, M.A. Emotion Recognition via Detrended Fluctuation Analysis and Fractal Dimensions. In Proceedings of the 2019 36th National Radio Science Conference (NRSC), Port Said, Egypt, 16–18 April 2019; pp. 200–208. [\[CrossRef\]](#)
33. Lozano-Monazor, E.; López, M.T.; Vigo-Bustos, F.; Fernández-Caballero, A. Facial expression recognition in ageing adults: From lab to ambient assisted living. *J. Ambient. Intell. Humaniz. Comput.* **2017**, *8*, 567–578. [\[CrossRef\]](#)
34. Ousmane, A.M.; Djara, T.; Vianou, A. Automatic recognition system of emotions expressed through the face using machine learning: Application to police interrogation simulation. In Proceedings of the 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, France, 24–26 April 2019; pp. 1–4. [\[CrossRef\]](#)
35. Zhang, K.; Zhang, H.; Li, S.; Yang, C.; Sun, L. The PMemo Dataset for Music Emotion Recognition. In Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, Yokohama, Japan, 11–14 June 2018; pp. 135–142. [\[CrossRef\]](#)
36. Gouverneur, P.; Jaworek-Korjakowska, J.; Köping, L.; Shirahama, K.; Kleczek, P.; Grzegorzec, M. Classification of Physiological Data for Emotion Recognition. In Proceedings of the Artificial Intelligence and Soft Computing, Zakopane, Poland, 11–15 June 2017; Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 619–627.

37. Rivera, H.; Valadão, C.; Caldeira, E.; Krishnan, S.; Bastos Filho, T.F. Development of a Toolkit for Online Analysis of Facial Emotion. In Proceedings of the XXVI Brazilian Congress on Biomedical Engineering, Armação de Buzios, RJ, Brazil, 21–25 October 2018; Costa-Felix, R., Machado, J.C., Alvarenga, A.V., Eds.; Springer: Singapore, 2019; pp. 619–625.
38. Bevilacqua, F.; Engström, H.; Backlund, P. Game-Calibrated and User-Tailored Remote Detection of Stress and Boredom in Games. *Sensors* **2019**, *19*, 2877. [\[CrossRef\]](#)
39. Hassani, S.; Bafadel, I.; Bekhatro, A.; Al Blooshi, E.; Ahmed, S.; Alahmad, M. Physiological signal-based emotion recognition system. In Proceedings of the 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), Salmabad, Bahrain, 29 November–1 December 2017; pp. 1–5. [\[CrossRef\]](#)
40. Li, B.J.; Bailenson, J.N.; Pines, A.; Greenleaf, W.J.; Williams, L.M. A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures. *Front. Psychol.* **2017**, *8*, 2116. [\[CrossRef\]](#)
41. Wei, Y.; Wu, Y.; Tudor, J. A real-time wearable emotion detection headband based on EEG measurement. *Sens. Actuators A Phys.* **2017**, *263*, 614–621. [\[CrossRef\]](#)
42. Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; Van Laerhoven, K. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction, ICMI '18, Boulder, CO, USA, 16–20 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 400–408. [\[CrossRef\]](#)
43. Carneiro, D.; Novais, P. Quantifying the effects of external factors on individual performance. *Future Gener. Comput. Syst.* **2017**, *66*, 171–186. [\[CrossRef\]](#)
44. Domínguez-Jiménez, J.; Campo-Landines, K.; Martínez-Santos, J.; Delahoz, E.; Contreras-Ortiz, S. A machine learning model for emotion recognition from physiological signals. *Biomed. Signal Process. Control* **2020**, *55*, 101646. [\[CrossRef\]](#)
45. Amira, T.; Dan, I.; Az-eddine, B.; Ngo, H.H.; Said, G.; Katarzyna, W. Monitoring chronic disease at home using connected devices. In Proceedings of the 2018 13th Annual Conference on System of Systems Engineering (SoSE), Paris, France, 19–22 June 2018; pp. 400–407. [\[CrossRef\]](#)
46. Cominelli, L.; Carbonaro, N.; Mazzei, D.; Garofalo, R.; Tognetti, A.; Rossi, D.D. A Multimodal Perception Framework for Users Emotional State Assessment in Social Robotics. *Future Internet* **2017**, *9*, 42. [\[CrossRef\]](#)
47. Liu, X.; Xie, L.; Wang, Z. Empathizing with emotional robot based on cognition reappraisal. *China Commun.* **2017**, *14*, 100–113. [\[CrossRef\]](#)
48. Navarro Tuch, S.; López-Aguilar, A.; Bustamante-Bello, R.; Molina, A.; Izquierdo-Reyes, J.; Curiel-Ramirez, L. Emotional domotics: A system and experimental model development for UX implementations. *Int. J. Interact. Des. Manuf. (IJIDeM)* **2019**, *13*, 1587–1601. [\[CrossRef\]](#)
49. Artífice, A.; Ferreira, F.; Marcelino-Jesus, E.; Sarraipa, J.; Jardim-Gonçalves, R. Student's Attention Improvement Supported by Physiological Measurements Analysis. In Proceedings of the Technological Innovation for Smart Systems, Costa de Caparica, Portugal, 3–5 May 2017; Camarinha-Matos, L.M., Parreira-Rocha, M., Ramezani, J., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 93–102.
50. Scioscia, F.; Ruta, M.; Di Sciascio, E. From Biosignals to Affective States: A Semantic Approach. In Proceedings of the 2018 2nd International Conference on Computational Biology and Bioinformatics, ICCBB 2018, Bari, Italy, 11–13 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 78–83. [\[CrossRef\]](#)
51. Forooghifar, F.; Aminifar, A.; Atienza Alonso, D. Self-Aware Wearable Systems in Epileptic Seizure Detection. In Proceedings of the 2018 21st Euromicro Conference on Digital System Design (DSD), Prague, Czech Republic, 29–31 August 2018; pp. 426–432.
52. Sopic, D.; Aminifar, A.; Aminifar, A.; Atienza, D. Real-Time Event-Driven Classification Technique for Early Detection and Prevention of Myocardial Infarction on Wearable Systems. *IEEE Trans. Biomed. Circuits Syst.* **2018**, *12*, 982–992. [\[CrossRef\]](#)
53. Zhang, Z.; Song, Y.; Cui, L.; Liu, X.; Zhu, T. Emotion recognition based on customized smart bracelet with built-in accelerometer. *PeerJ* **2016**, *4*, e2258. [\[CrossRef\]](#)
54. Hovsepian, K.; Al'Absi, M.; Ertin, E.; Kamarck, T.; Nakajima, M.; Kumar, S. cStress: Towards a gold standard for continuous stress assessment in the mobile environment. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015; pp. 493–504.
55. Healey, J.A.; Picard, R.W. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [\[CrossRef\]](#)
56. Vyzas, E. Recognition of Emotional and Cognitive States Using Physiological Data. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1999.
57. Mijic, I.; Sarlija, M.; Petrinovic, D. MMOD COG: A Database for Multimodal Cognitive Load Classification. In Proceedings of the 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 23–25 September 2019; pp. 15–20.
58. Koldijk, S.; Sappelli, M.; Verberne, S.; Neerincx, M.; Kraaij, W. The SWELL Knowledge Work Dataset for Stress and User Modeling Research. In Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, 12–16 November 2014. [\[CrossRef\]](#)
59. Nkurikiyeyezu, K.; Yokokubo, A.; Lopez, G. The Influence of Person-Specific Biometrics in Improving Generic Stress Predictive Models. *arXiv* **2019**, arXiv:1910.01770.

60. Quiroz, J.C.; Yong, M.H.; Geangu, E. Emotion-recognition using smart watch accelerometer data: Preliminary findings. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, Maui, HI, USA, 11–15 September 2017; pp. 805–812.
61. Delgado Reyes, A.; Parra, T.; Lopez, J. Realidad virtual: Evaluacion e Intervencion en el Trastorno del Espectro Autista. *Rev. Electrónica Psicol. Iztaclala* **2020**, *23*, 369–399.
62. Fluja, J.; Castañeda, D.; Becerra, I. Promoting Emotional Well-being in Hospitalized Children and Adolescents With Virtual Reality: Usability and Acceptability of a Randomized Controlled Trial. *CIN Comput. Informat. Nurs.* **2019**, *38*, 1. [[CrossRef](#)]
63. Haufe, S.; Treder, M.S.; Gugler, M.F.; Sagebaum, M.; Curio, G.; Blankertz, B. EEG potentials predict upcoming emergency brakings during simulated driving. *J. Neural Eng.* **2011**, *8*, 056001. [[CrossRef](#)] [[PubMed](#)]
64. Hallvig, D.; Anund, A.; Fors, C.; Kecklund, G.; Karlsson, J.G.; Wahde, M.; Åkerstedt, T. Sleepy driving on the real road and in the simulator—A comparison. *Accid. Anal. Prev.* **2013**, *50*, 44–50. [[CrossRef](#)] [[PubMed](#)]
65. Jagannath, M.; Balasubramanian, V. Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator. *Appl. Ergon.* **2014**, *45*, 1140–1147. [[CrossRef](#)] [[PubMed](#)]
66. Kanjo, E.; Younis, E.M.; Sherkat, N. Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach. *Inf. Fusion* **2018**, *40*, 18–31. [[CrossRef](#)]
67. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [[CrossRef](#)]
68. Nalepa, G.J.; Kutt, K.; Bobek, S. Mobile platform for affective context-aware systems. *Future Gener. Comput. Syst.* **2019**, *92*, 490–503. [[CrossRef](#)]
69. Santamaria-Granados, L.; Munoz-Organero, M.; Ramirez-González, G.; Abdulhay, E.; Arunkumar, N. Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS). *IEEE Access* **2019**, *7*, 57–67. [[CrossRef](#)]
70. Chiang, H.S. ECG-based Mental Stress Assessment Using Fuzzy Computing and Associative Petri Net. *J. Med. Biol. Eng.* **2015**, *35*, 833–844. [[CrossRef](#)]
71. Saganowski, S.; Dutkowiak, A.; Dziadek, A.; Dziezyc, M.; Komoszyńska, J.; Michalska, W.; Polák, A.; Ujma, M.; Kazienko, P. Emotion Recognition Using Wearables: A Systematic Literature Review Work in progress. *arXiv* **2019**, arXiv:1912.10528.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.